Chapter XVI

Spatial Reasoning for Human-Robot Teams

David J. Bruemmer, Idaho National Laboratory, USA
Douglas A. Few, Idaho National Laboratory, USA
Curtis W. Nielsen, Idaho National Laboratory, USA

Abstract

This chapter presents research designed to study and improve an operator's ability to navigate or teleoperate a robot that is distant from the operator through the use of a robot intelligence architecture and a virtual 3D interface. To validate the use of the robot intelligence architecture and the 3D interface, four user-studies are presented that compare intelligence modes and interface designs in navigation and exploration tasks. Results from the user studies suggest that performance is improved when the robot assumes some of the navigational responsibilities or the interface presents spatial information as it relates to the pose of the robot in the remote environment. The authors hope that understanding the roles of intelligence and interface design when operating a remote robot will lead to improved human-robot teams that are useful in a variety of tasks.

Introduction

Robots have been used in a variety of settings where human access is difficult, impractical, or dangerous. These settings include search and rescue, space exploration, toxic site cleanup, reconnaissance, patrols, and many others (Murphy, 2004). Often, when a robot is used in one of these conditions, the robot is distant from the operator; this is referred to as *teleoperation*. Ideally, robots could be a useful member of a team because they could be used to accomplish tasks that might be too difficult or impractical for a human to perform.

The potential, however, for humans and robots to work as an effective team is limited by the lack of an appropriate means for the operator to visualize the remote environment and how the robot fits within the environment. As an example, several recent research efforts have investigated the human-robot interaction challenges associated with real-world operations including search and rescue and remote characterization of high-radiation environments (Burke, Murphy, Coovert, & Riddle, 2004; Casper & Murphy, 2003; Murphy, 2004; Yanco, Drury, & Scholtz, 2004a). Across these disparate domains, researchers have noted that it is difficult for operators to navigate a remote robot due to difficulty and error in operator understanding of the robot's position and/or perspective within the remote environment.

A primary reason for the difficulty in remote robot teleoperation is that for the overwhelming majority of robotic operations, video remains the primary means of providing information from the remote environment to the operator (Burke, Murphy, Rogers, Lumelsky, & Scholtz, 2004a). Woods, Tittle, Feil, and Roesler (2004) describe the process of using video to navigate a robot as attempting to drive while looking through a "soda straw" because of the limited angular view associated with the camera (Woods et al., 2004). The limited angular view of the camera presents problems for robot teleoperation because obstacles outside of the field of view of the camera still pose navigational threats to the robot even though they are not visible to the operator.

To alleviate navigational threats to the robot, current research at the Idaho National Laboratory (INL) is aimed at providing tools that support mixed-initiative control where humans and robots are able to make decisions and take initiative to accomplish a task. The goal is to create a set of capabilities that permit robots to be viewed as trusted teammates rather than passive tools. If this is to happen, the robot as well as the human must be enabled to reason spatially about the task and environment. Furthermore, true teamwork requires a shared understanding of the environment and task between team members in order to understand each others' intentions (Dennett, 1981). The lack of an effective shared understanding has been a significant impediment to having humans and intelligent robots work together.

In response to this challenge, the INL has developed a mixed-initiative robot control architecture that provides a framework for robot intelligence, environment modeling, and information sharing. In order to support a shared understanding of the environment and task between robotic and human team members, a virtual three dimensional 3D interface was developed through collaboration with researchers at Brigham Young University (BYU). The combination of the virtual 3D interface and intelligence on the robot can be used to improve the human's and robot's ability to reason spatially about the environment by presenting a shared understanding of the environment. The technology used to achieve the shared understanding of the environment is presented next.

System Design

Through the Office of the Secretary of Defense (OSD) Joint Robotics Program (JRP), the Space and Naval Warfare Systems Center (SPAWAR) at San Diego and the INL have worked together to develop, mature, and integrate promising robotics technologies from throughout the robotics community including components for perception, communication, behavior, and world modeling. One of the results of this collaboration is the development of the INL Robot Intelligence Architecture which is currently used to unite selected components into a behavior-based intelligence kernel that can be transferred to a variety of fieldable, unmanned ground vehicle systems.

The robot intelligence architecture is the product of an iterative development cycle where behaviors have been evaluated in the hands of users, modified, and tested again. In fact, many of the strategies and interface components that originally seemed elegant from a conceptual standpoint, proved to be frustrating for users. For example, during a preliminary experiment that evaluated robot intelligence, but provided minimal spatial reasoning tools to the operator, it was noted that although most participants felt a high level of control, some participants indicated that they were confused by the robot behaviors (Marble, Bruemmer, & Few, 2003). In particular, the automatic initiation of robot behaviors to get the robot out of a narrow hallway led to operator confusion and a fight for control between the robot and human, because operators thought the robot could go through the hallway but the robot sensors indicated that it would not fit. The lack of adequate spatial representation tools prevented the operator from realizing this fact and, consequently, the human and robot engaged in a fight for control of the robot's movements. Findings such as these serve to motivate improvements to the robot intelligence architecture and the development of interface components that could support spatial reasoning.

Currently the robot intelligence architecture is divided into four modes of control (Tele, Safe, Shared, and Autonomous) affording the robot different types of behavior and levels of autonomy (Marble, Bruemmer, & Few, 2003; Marble, Bruemmer, Few, & Dudenhoeffer, 2004). The modes of autonomy in the robot intelligence architecture include:

1. **Tele mode** is a fully-manual mode of operation, in which the operator must manually control all robot movement.
2. **Safe mode** is similar to Tele Mode, in that robot movement is dependent on manual control. However, in safe mode, the robot is equipped with a level of initiative that prevents the operator from colliding with obstacles.
3. In **Shared mode**, the robot can relieve the operator from the burden of direct control, using reactive navigation to find a path based on perception of the environment. Shared Mode provides for a dynamic allocation of roles and responsibilities. The robot accepts varying levels of operator intervention and supports dialogue through the use of a finite number of scripted suggestions (e.g., "Path blocked! Continue left or right?") and other text messages that appear in a text box within the graphical interface.
4. **Autonomous mode** consists of a series of high-level tasks such as patrol, search a region, follow a path, or go to a place. In Autonomous Mode, the only user intervention occurs on the tasking level; the robot itself manages all navigational decision-making.

To investigate the challenges of sharing control of the robot between the robot and the operator, the experiments reported in this research focus on the middle ground that falls between teleoperation and full robotic autonomy (i.e., safe mode and shared mode). Although the experiments restricted each participant to only one level of control, normal operation would permit the user to switch between all four modes of autonomy as the task constraints, human needs, and robot capabilities change. As an example, tele mode could be useful to push open a door or shift a chair out of the way, whereas autonomous mode could be used to reduce human workload or in an area where communications to and from the robot are sporadic.

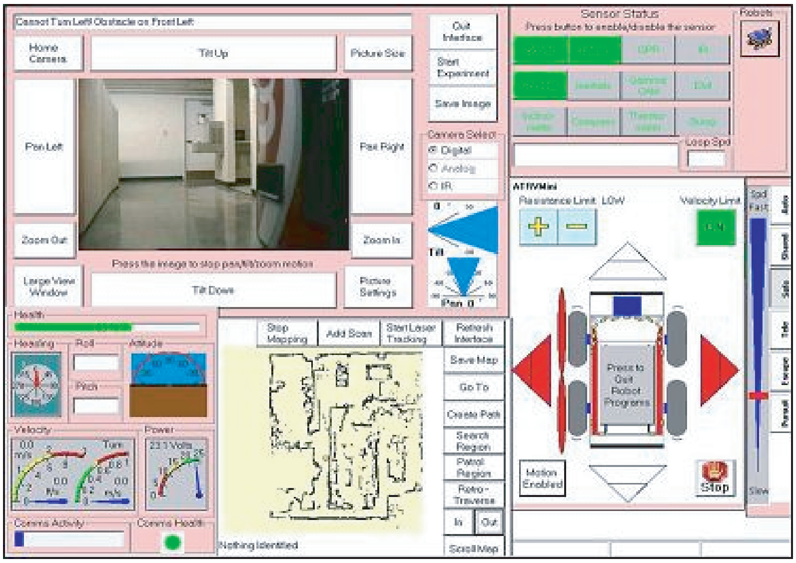
In order to protect the robot from collisions with obstacles in robot control modes that have some robot autonomy (safe, shared, autonomous), a *guarded motion* behavior based on a technique described by Pacis, Everett, Farrington, and Bruemmer (2004) is implemented. In response to laser and sonar range sensing of nearby obstacles, the guarded motion behavior scales down the robot's velocity using an event horizon calculation, which measures the maximum speed at which the robot can safely travel in order to come to a stop approximately two inches from an obstacle. By scaling down the speed in small increments, it is possible to insure that regardless of the commanded translational or rotational velocity, guarded motion will stop the robot at a consistent distance from an obstacle. This approach provides predictability and ensures minimal interference with the operator's control of the

Figure 1. The robots used for Experiments 1 – 4

vehicle. If the robot is being driven near an obstacle rather than directly towards it, guarded motion will not stop the robot, but may slow its speed according to the event horizon calculation.

In order for the robot to be navigated successfully, spatial information of the environment must be available. The robot intelligence architecture gathers spatial

Figure 2. The standard interface



information from the environment with a laser range finder. Information from laser scans is combined into a map of the environment using a technique developed at the Stanford Research Institute (SRI) called consistent pose estimation (CPE) (Gutman & Konolige. 1999; Konolige, 2004). This map-building algorithm is designed to build an occupancy-grid based map of the robot's environment as the robot explores the environment (Elfes, 1987; Moravec, 1988). The mapping algorithm is particularly useful because it provides an accurate spatial representation of complex environments that are previously unknown to the robot or the operator.

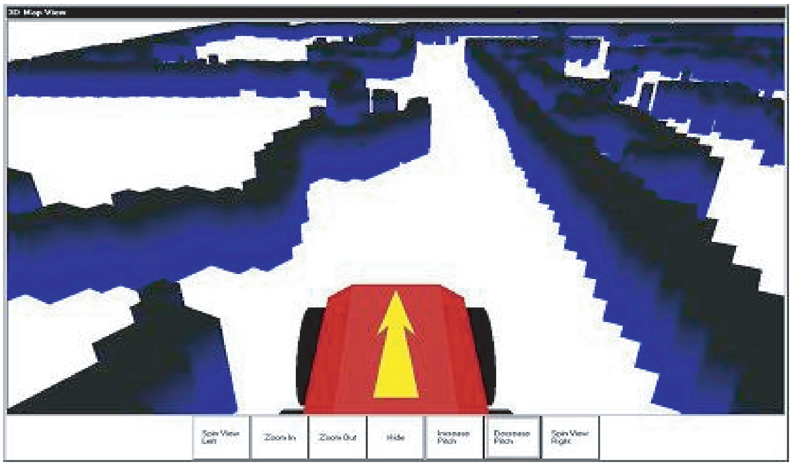
Since no single robot platform is appropriate for all tasks, the INL robot intelligence architecture can port to a variety of robot geometries and sensor suites and is currently in use as a standard by several research teams throughout the human-robot interaction (HRI) community. Experiments presented later in this paper were performed with an iRobot "ATRV mini" or an iRobot "ATRV Jr" shown in Figure 1. On each robot, the intelligence architecture utilizes a variety of sensor information including inertial sensors, compass, wheel encoders, laser, computer vision, thermal camera, infrared break beams, tilt sensors, bump sensors, sonar, and ultrasonic sensors.

The default configuration of the interface used to interact with the robot consists of a single touch screen display containing five re-sizeable windows as shown in Figure 2 (Bruemmer et al., 2005). The upper left-hand window on the screen contains a video feed from the robot as well as controls for panning, tilting, and zooming the camera. Frame size, frame rate, and compression settings can be accessed from a sub-window, but are held constant throughout the experiments reported here.

The upper right-hand window contains sensor status indicators and controls that allow the operator to monitor and configure the robot's sensor suite as needed. The lower right-hand window pertains to movement within the local environment and provides indications of robot velocity, obstructions, resistance to motion, and feedback from contact sensors. The interface indicates blockages that impede motion in a given direction as red ovals next to the iconographic representation of the robot wheels (lower right of Figure 2). The current snapshot of the interface indicates that movement right and left is not possible because of an object close to the wheels on the left side of the robot. These indicators are designed to inform the operator as to why the robot has overridden a movement command. Since the visual indications can sometimes be overlooked, a force feedback joystick is also implemented to resist movement in the blocked direction. The joystick vibrates if the user continues to command movement in a direction already indicated as blocked. At the far right of the window the user can select between different levels of robot autonomy.

The lower central window displays the map of the environment as it is discovered by the robot and allows the user to initiate a number of waypoint-based autonomous behaviors such as *search region, patrol region*, *create a path*, or *go to a place*. Additionally, the map can be moved and zoomed in and out to provide a desired perspective. The lower left-hand window contains information about the robot's operational status such as communication activity, power, and the robot's pitch and roll.

Figure 3. The virtual 3D interface



The virtual three-dimensional (3D) display (Figure 3) was designed to support the operator's awareness of the spatial information in the robot's environment and show the information related to the robot's current pose within the environment. The virtual 3D component has been developed by melding technologies from the INL (Bruemmer et al., 2005), Brigham Young University (BYU) (Nielsen, Goodrich, & Crandall, 2004; Ricks, Nielsen, & Goodrich, 2004), and the Stanford Research Institute (SRI) (Gutman & Konolige, 1999; Konolige, 2004). The 3D display is not based on true 3D range sensing, but rather by extruding the 2D map built by the robot into a 3D perspective. The map information in the 3D interface and the standard interface both originate from the map-building algorithm on the robot. The only difference is the manner in which the information is presented to the operator.

The map information produces the basis for the 3D representation that includes obstacles and other semantic entities that are of significance to the operator such as start location, labels, and waypoints. These items can be inserted by the robot to indicate percepts and intentions or by the human to identify and classify targets in the environment. Also, the user is able to add, verify, remove, or annotate semantic entities displayed within the map. Collaborative construction of the map enhances each individual team member's understanding of the environment and provides a basis for the human-robot team to "communicate" naturally about the environment through the visualization of relevant spatial information.

In the 3D interface, the operator may also insert translucent still images excerpted from the robot video, which are overlaid onto the corresponding area of the 3D map display, providing a means to fuse real video information with the virtual representation of the environment (Nielsen, Goodrich, & Crandall, 2004). By changing the virtual display's zoom, pitch, and yaw, it is possible to move the virtual perspective of the robot and environment from an egocentric perspective (i.e., looking out from the robot), to a fully exocentric view where the entire environment (map) can be seen at once.

The experiments presented next utilize the robot intelligence architecture as described in this section to compare the safe and shared modes of robot control along with the use of the standard interface and the 3D interface in navigation and exploration tasks.

Experiment 1

The first experiment was intended to: (a) show that the behaviors on board the robot (e.g. guarded motion and autonomous navigation) were useful in an exploration task, and (b) to compare the safe and shared modes of autonomy in an exploration task. It was hypothesized that participants would perform better with the shared autonomy mode than with the safe autonomy mode.

For this experiment a 20’ × 30’ maze environment was created using conventional office dividers and cylindrical pylons. Participants controlled the robot from a remote station where the robot environment was not visible. Five objects of interest (two mannequins, a stuffed dog, a disabled robot, and a small simulated explosive device) were placed throughout the arena in locations that remained fixed for all participants. The placement of these items further complicated the navigation task since operators were told not to drive into or over the objects. Moreover, certain objects remained hidden except from certain vantage points so the operator was required to maneuver the robot in order to see all the objects of interest.

Each participant was given 60 seconds to locate as many of the five items in the search area as possible using the standard interface and either the *safe* or the *shared* autonomy mode. Prior to the experiment, participants were instructed on the use of the joystick for controlling the robot and the camera on the robot (e.g. pan, tilt, and zoom), but were given no opportunity to practice controlling the robot until the experiment began. Operators with the *safe* autonomy mode were reminded that the robot would take initiative to avoid collisions but that they (the operators) should seek to avoid collisions as well. Operators with the *shared* autonomy mode were reminded to let the robot do most of the driving, but that if they wanted to redirect the robot, it would temporarily yield control to their joystick commands.

There were 107 participants drawn as volunteers from attendees of the 2003 INL annual science and engineering exposition at the Museum of Idaho in Idaho Falls. The participants consisted of 46 females and 61 males, ranging in age from 3 to 78 years old, with a mean age of 14. Participants were asked demographic questions including their age and gender, and whether they had experience in remote systems operation. It was determined by self-report that none of the participants had experience remotely controlling robots, or had knowledge of or access to the remote environment. Furthermore, none had prior experience with or knowledge of the interface or robot control system; therefore, it was determined that all of the participants could be regarded as novice users. Participants were assigned to either the *shared* or *safe* autonomy modes alternately based on their sequence in participation.

On average, participants who used the robot's *shared* autonomy mode found an average of 2.87 objects while those who used the *safe* autonomy mode found an average of 2.35 objects (Bruemmer et al., 2005). Comparisons between different age groups and gender were analyzed, but a significant difference in the number of items found did not exist based on age or gender. Although this experiment was not intended to support a careful comparison of age and gender groupings, it does support the claim that the interface allowed a wide variety of participants to find objects successfully. Participants were able to find objects successfully in both safe mode and shared mode, indicating that both the guarded motion used in safe mode and the autonomous navigation behaviors used in shared mode were usable by participants. Across all age and gender groupings, performance was better in shared mode than in safe mode, providing evidence that the robot's ability to navigate the environment can actually exceed the ability of a human operator. The performance benefit experienced by allowing the robot to navigate suggests the potential to use robot initiative and autonomy not only as a last resort (i.e., when communication fails or operator workload increases), but as a basis for collaborative interaction between a human and a robot.

Taken on its own, this first study demonstrates the utility of robot autonomy, but leaves many questions to be answered by further experiments. The first experiment did not look beyond overall performance (as measured by items found) to discern the reasons for the observed difference in performance between safe mode and shared mode. In response to this limitation, it was determined that the next experiments should empirically measure differences in operator workload, operator error, and operator confusion in order to provide deeper insight. Additionally, this experiment utilized a relatively small search environment. Areas of the environment required careful maneuvering, but the task was not designed to reward path planning or strategy. Future experiments address this question by using larger environments that require some path-planning and strategy to explore the environment efficiently.

Experiment 1 also raised the question of how useful the streaming video provided by the interface actually was to users when navigating the robot. In tight spaces where spatial information is important to prevent collisions, participants often found the entire visual field filled by an immediate obstacle, thereby diminishing the usefulness of the video for navigation. Furthermore, video information fails to illustrate obstacles outside of the current visual field, which makes it difficult for the operator to remember the location of obstacles on the sides of the robot. One hypothesis was that in such instances video promoted a false sense of spatial awareness and led to operator confusion. As an example, consider the common scenario of a robot approaching an open doorway in safe mode. The door frame disappears from the video feed before the robot has reached the doorway. However, the operator, already viewing video information from the next room, may believe that the robot is already through the door. To prevent a collision with the doorframe, the robot may stop and refuse to move forward. Although the robot communicates that it is blocked in front, the user may be confused by the lack of obstacles in the visual feed. Put simply, the default interface used in Experiment 1 did not provide the operator with an adequate representation of the spatial information around the robot. Experiment 2 was designed to explore the use of a new interface component intended to better support an operator's understanding of the spatial information around the robot.

Experiment 2

Observations from Experiment 1 suggest that video may not provide an adequate perspective of the remote environment, let alone means for the operator to predict robot behavior or understand the robot's intentions. However, humans are visual and prefer pictures and diagrams when attempting to understand or communicate (Pashler, 1990). In order to address the *human-robot-interaction* (HRI) limitations observed in Experiment 1, some means were required to support collaborative understanding and yet take advantage of the functional utility associated with visual representation. In addition to these human factors, there were also significant engineering reasons for assessing alternatives to video presentation of the remote environment. In particular, video demands high-bandwidth, continuous communication, and is therefore ill-suited for many of the very environments where robots could be most useful. Except for short ranges, transmission of high-bandwidth video is only possible when line of sight can be maintained either with a satellite or another radio antenna. For instance, high-bandwidth video cannot be transmitted through layers of concrete and rebar, making it inappropriate for urban terrain or urban search and rescue. Likewise, forest and jungle canopy precludes reliable transmission of video.

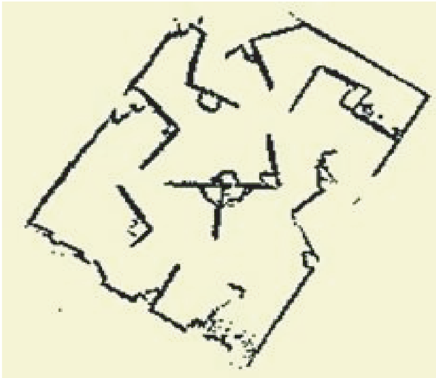
In response to these human and engineering factors, collaboration between the INL and Brigham Young University (BYU) was used to develop a new 3D interface component that could provide a better perspective of the spatial structure of the environment around the robot. The improved presentation of the spatial information may help the operator gain insight into the reason for robot initiative and diminish the likelihood of operator confusion. The purpose of this experiment is to assess the effectiveness of the 3D interface in a spatial exploration task where the operators were to use the robot to construct a map of an environment. The hypothesis was that the 3D interface without video would support the operator in the spatial exploration task better than the standard interface with video.

Figure 4. A partial view of the arena built at the St. Louis Science Center

The experiment was performed over a seven-day period within the St. Louis Science Center in 2004 and utilized 64 visitors who volunteered to take part in the experiment. The majority of participants were high school students from schools in the St. Louis area. These students were not pre-selected, but rather volunteered to take part in the study while visiting the Science Center. As before, the experiment was set up as a remote deployment such that the operator control station was located several stories above the robot arena so that the operator could not see the robot or the robot's environment. The arena was built by the production staff of the Science Center and contained artificial rocks, artificial trees, mannequins, and plywood dividers to create a maze environment (Figure 4).

Due to the distance and physical occlusions separating the control station from the actual robot environment, analog video was not possible. Instead, state-of-the-art video compression was used to digitize the analog video into a motion JPEG (MJPEG) format and wirelessly transmit from the robot to a nearby access point connected to the building's network. The building's wired network was then used to transfer the video data two stories up to the operator. Exploiting the wired infrastructure in place throughout the building made it possible to provide continuous, reliable video at a high frame rate. The presentation speed and resolution of this video exceeded that possible through an entirely wireless data link. This configuration ensured that the comparison between video and the 3D map display was not merely a function of current communication bandwidth constraints, but rather an investigation of the fundamental differences between an interface based primarily on viewing raw video and one which presented the environment and obstacles as they relate to the robot's pose.

Figure 5. A near-complete map built up by one of the participants



Before the experiment, each participant was given basic instructions on how to use the interface, and no participant was permitted to drive the robot until the start of the trial run. Participants only used the safe mode autonomy level in order to simplify the comparison of performance between the two interfaces. Participants were assigned to alternating display conditions (standard interface with video, standard interface with the 3D window in place of the video) in order to ensure equal numbers of participants in each condition and no participant was allowed to operate the robot in more than one trial. A time limit of three minutes was set in place to help insure that the measured performance was a function of the interface presentation rather than a function of operator interest or time spent on the task.

At the beginning of each experiment, the map built by the previous participant was erased by restarting the map-building algorithm on the robot. Each participant was then instructed to drive the robot around the environment in order to build as large a map as possible as quickly as possible. All participants were given access to the same 2D map component (Figure 5) within which the robot presents the map that it builds as it explores new territory. Exactly half of the participants used the standard interface and were able to see both the 2D map and the video module. The other half of participants used the same interface except that the 3D interface module entirely occluded the video module.

During each trial, the interface stored a variety of useful information about the participant's interactions with the robot. Joystick bandwidth was recorded as the number of messages sent from the joystick indicating a change of more than 10% in the position of the stick. This information is used as an indirect measure of workload (Clarke, Yen, Kondraske, Khoury, & Maxwell, 1991; Khoury & Kondraske, 1991). The interface also recorded the number of joystick vibrations caused by human navigational error. The map produced by the robot for each experiment was also saved in order to assess performance based on coverage of the environment.

This approach provided a reasonable assessment of the operator's ability to explore the environment in the time available. Immediately after completing a trial, each participant was asked to rank on a scale of 1 to 10 how "in control" they felt during the operation, where 1 signified "The robot did nothing that I wanted it to do" and 10 signified, "The robot did everything I wanted it to do."

In the three minutes provided, 80% of the participants explored over half of the total environment. One person, a 3D display participant, was able to build the entire map in the allotted 3 minutes. As described above, task performance was calculated by comparing the map generated during the exploration task with the complete map of the task environment. This comparison showed no significant difference between the use of the video module and the 3D module. Using joystick bandwidth as an indication of human workload and joystick vibration as a metric for human navigational error, analysis shows that operators using the virtual 3D display worked less and demonstrated fewer instances of navigational error. On average, the joystick bandwidth for participants using the 3D module was 1,057 messages from the interface to the robot, compared to 1,229 average messages for operators using the video module. Further, there were, on average, 11.00 instances of navigational error with the 3D module and 14.29 instances with the video module as measured by joystick vibrations (Bruemmer et al., 2005).

In addition to reduced workload and fewer navigational errors, use of the virtual 3D display slightly increased the operator's subjective "feeling of control" while operating the robot. The average feeling of control for the 3D display was 7.219 compared with an average of 7.059 for the video.

The second experiment provided initial evidence that the virtual 3D perspective of the robot's environment supported an operator's ability to reason spatially about the task and environment better than streaming video information. Results suggest that although there was no significant change in performance (as measured by the percentage of the map discovered), there was reduced operator workload, less navigational error, and a slightly improved sense of control.

One motivation for the development of the virtual 3D display had been to promote a shared understanding between the robot and the operator of the task and the robot's environment. To assess the effectiveness of the virtual 3D display in this regard, it is useful to consider that a decrease in joystick vibrations not only represents a reduction in operator navigational error, but also a reduction in the instances where the operator failed to understand the reason the robot took initiative to protect itself. Recall that the joystick vibrates only if the operator commands movement in a direction in which the robot has already recognized an obstacle and taken initiative to prevent a collision. These results indicate progress towards the goal of providing a representation that supports spatial reasoning and a shared understanding of the environment. More broadly, these results provide evidence that it may be possible to support navigational needs of human operators without using video. This finding provides an important counterpoint to opinion within the field of human-robot interaction that reliable, continuous video is essential for remote navigation (Baker, Casey, Keyes, & Yanco, 2004).

From an engineering perspective, this experiment shows that it is possible to have a robot build a map of the robot's environment as the environment is explored and communicate the map back to a remote user fast enough to support real-time robot navigation by the operator. The significance of this result to the area of remote systems can be seen most clearly when one considers the reduction in communication bandwidth made possible by using the 3D map display. Whereas the video alone requires 3,000,000 bits per second (bps), the total interface bandwidth with the virtual 3D interface was only 64,000 bps. This bandwidth savings allows control to extend into new domains using data transmission methods that can be used in underground bunkers, caves, nuclear reactors, and urban search and rescue sites where it is often impossible to maintain a video feed.

Despite the fact that the human-robot team can function effectively without video, there is no reason to disregard the potential benefits of video in those instances when video is available. Experience with operators and subject area experts from energy, defense, and emergency management contexts indicate that operators expect and can exploit video in remarkable ways (Casper & Murphy, 2003; Marble, Bruemmer, & Few, 2003; Yanco & Drury, 2004;). Many applications require the human to play a role in visual search and detection. Although this experiment suggests that video could be replaced with the 3D representation, the optimal interface will likely provide a dynamic balance between the video and virtual displays.

Experiment 3

The second experiment showed that the virtual 3D display could support the operator's comprehension of the spatial information regarding the task and environment. The question still remains, however, as to whether the use of the virtual 3D display could be improved with shared control where the human and robot engage in a dynamic sharing of roles and responsibilities. A previous usability study by Marble et al. (2003) showed that shared mode offered the greatest potential for operator confusion and frustration. Consequently, it was hypothesized that Shared Mode might provide the greatest potential for the virtual 3D display to reduce navigational error and operator workload.

One goal of this experiment is to compare the safe and shared autonomy modes when the 3D interface is used. Another goal is to show that the benefits of sharing control between the human and operator observed in the first experiment are not merely due to the high cognitive workload placed on the operator when using the standard interface, but are related to the robot's ability to navigate itself through the environment. The typical assumption found in the literature is that robot autonomy trails behind human performance, but may be useful when the human's ability to spatially reason about the task and environment is encumbered (i.e., operator workload increases, communications fail, or map and position accuracy begins to degrade) (Goodrich, Olsen Jr., Crandall, & Plamer, 2001; Nielsen, Goodrich, & Crandall, 2003; Trouvain, Wolf, & Schneider, 2003). It was hoped that this experiment could provide evidence that the robot's ability to reason about the environment can improve performance even when the operator's ability to reason spatially is unhindered (i.e., data link connectivity is maintained, human workload is minimal, and mapping and localization is reliable).

The task for this experiment is similar to that of Experiment 1 where the participants were asked to find as many items of interest as possible. For this experiment, however, in order to minimize individual human workload, the control task was separated into specific operator functions, namely navigation, driving, and operation of a pan, tilt, and zoom camera. Instead of using only individuals, groups of participants were assigned roles where members had responsibility over one aspect of the robot control. In addition to minimizing individual human workload, an added benefit of assigning different roles was that it afforded an opportunity to observe the exchange of information between team members in different roles. In fact, it became very clear that operators in different roles require different perspectives. For example, the navigation or planning role requires an exocentric display where the operator can see the entire environment while the driving role requires an egocentric perspective so the operator can visualize the robot's situation in the environment. As Scholtz (2002) points out, the roles of human operators do not remain static, and interfaces should be able to adapt accordingly.

This experiment included 120 volunteers grouped into teams of six members. The participating teams consisted of one team of teachers, three teams of eighth grade students, and the remainder of the teams being drawn from local high schools. Participants were recruited from a solicitation of local schools through the St. Louis Science Center's outreach program. Participants knew and selected the other people in their team prior to participation in the experiment. Age and gender were not recorded due to the fact that most participants were of similar age and the fact that gender was mixed for each team.

The experiment was run over seven days at the St. Louis Science Center in 2004. Teams of participants were assigned to alternating conditions so as to ensure equal numbers of teams in each condition. No participant was allowed to take part on more than one team. As in the previous experiment, the robot was located in the lower level of the Science Center, while the control center was located on the top level. This experiment used the same environment as was used in Experiment 2 with the same lighting and placement of obstacles. Three mannequins were placed in locations in the environment designed to force teams to coordinate their information in order to discover aspects regarding each particular mannequin's location. The mannequins remained in place throughout the entire experiment. An equal number of teams used the Shared and Safe modes of autonomy while controlling the robot. The interface components were divided across three separate stations, each with its own monitor and input devices. No interface component was visible at more than one control station. Two participants manned each station resulting in a total of six people dedicated to robotic system control. The stations were arranged in an arc such that the participants at each station could communicate easily with the others, but could not see the other displays.

The first control station was dedicated to the application payload, which in this case was a pan, tilt, and zoom camera. Using a joystick that allowed operation of the various camera controls, the application payload participants used the visual feedback from the robot to seek out the three mannequins and to provide navigational advice. The second control station was dedicated to driving the robot. Participants were permitted to see the virtual 3D interface along with the local environment window, the sensor status window, and the robot state window from the standard interface (Figure 2). Primarily, the operators at the driving station used the virtual 3D display, but were constrained to an egocentric perspective which precluded a global view of the environment. The final station was the navigation station where participants had access to the 2D map being built as the robot traveled through its environment. This gave them a bird's eye view of the environment and the robot's position in it. Additionally, participants at the navigation station were given a hard-copy of a map showing the locations of the three mannequins. Having two participants at each station was not necessary, but ensured that workload was minimal. Task completion required the three groups to self-organize in order to arrive at and gain a visual lock on all three of the mannequins as quickly as possible.

On average, less time was required to find the three mannequins for the teams using the Shared robot autonomy mode. The average completion time for Shared Mode teams was 467 seconds compared to an average completion time of 641 seconds for the safe mode teams. Safe mode teams also demonstrated a greater workload, as measured by joystick movement, than that of their Shared Mode counterparts. Safe mode teams made, on average, 2,744 significant joystick movements compared to an average of 1,725 significant joystick movements for shared mode teams. Using joystick vibration as a metric for human navigational error shows that safe mode teams made 25.1 errors on average compared to 16.8 errors for the Shared Mode teams (Bruemmer et al., 2005).

As with the first experiment, participants using the shared mode experienced increased performance efficiency when compared to their safe mode counterparts. The results from Experiment 3 show that with a representation that supports the human and robot's ability to reason spatially, performance can be significantly improved by sharing control between humans and a robot. Moreover, it shows that reducing the workload placed on the human driver and increasing the importance of strategy and intelligence does not diminish the performance benefits of sharing control between human and robot team members.

Previous research has shown that effective teams utilize a shared mental model of the task and current situation (Cooke, Salas, Cannon-Bowers, & Stout, 2000; Yen, Yin, Ioerger, Miller, Xu, & Volz, 2001). Similarly, the findings from Experiment 3 suggest that in order to fully realize the benefits of sharing control between human and robot team members, it is advantageous to provide a shared model of the environment. Unlike most interfaces for remotely controlling a mobile robot, the virtual 3D display presents information from the robot's environment from a perspective that helps the operator perceive and comprehend the spatial information around the robot. Improved comprehension of the robot's environment makes it easier for the operator to predict robot behavior and understand occasions of robot initiative (Endsley, 1988).

In many operational scenarios, it is not only possible, but probable that the roles of driving, navigating, and operating the application payload will be distributed among multiple human operators. Several researchers have pointed out the high cognitive burden associated with remote deployment of mobile robots and have argued that effective control requires multiple human operators (Burke et al., 2004; Casper & Murphy, 2003; Murphy, 2004). Although detailed analysis of these different roles (i.e., driver, navigator, payload operator) was beyond the scope of this experiment, anecdotal observations (recorded during and after the experiment) suggest interesting areas for further investigation. One observation was that just as performance can be degraded by a fight for control between the driver and robot; there is also the potential for similar conflicts between human operators primarily because they visualize the information differently. Their ability to reason spatially about the task and environment is dependent on the different perspectives associated with their roles. Further experimentation will be necessary to characterize the reasons for these choices and quantify their effect on team performance. One explanation found in the literature is that team success depends on the ability of each team member to understand the perspective of other members (Yen et al, 2001). If this is true, the most effective human-robot teams will be those that utilize a collaborative model of the environment and task. Such research questions provide a fertile ground for further experimentation into the challenges of providing shared representation, not only between human and robot, but also between humans.

Experiment 4

Experiments 1-3 showed that the INL control architecture, including the 3D interface and the robot intelligence architecture could reduce reliance on continuous video, increase overall task efficiency, and reduce operator error and workload. However, it is unclear what role the perspective of the virtual 3D environment had in bringing about these benefits. It is possible that the benefits due to the 3D perspective are largely due to the simplification brought through the abstraction process. However, it is also possible that the main benefit of the 3D display is that it provides a perspective that illustrates more of the spatial information near the robot and is, therefore, more useful for navigation and exploration tasks than the video display typically used in remote robot operation (teleoperation). The purpose of this study is to investigate the role of perspective in the 3D interface in terms of operator error, workload, and overall task efficiency.

This experiment included 216 participants drawn at random from attendees of the INL's 2004 annual community exposition. The participants consisted of 61 females and 155 males, ranging in age from 3 to 70 years old, with a mean age of 12. The robot used for this study was an ATRVmini designed by IRobot. Participants were assigned the task of discovering the physical structure of the environment using the Safe autonomy mode on the robot and the 3D interface which was populated by the map as the robot was navigated through the environment.

To test the role of perspective in the 3D interface, each volunteer was assigned one of four different perspectives (first person, close, elevated, and far). The *first person* perspective places the camera inside the robot, so the view is what it would be if the participant was sitting in the robot. It is similar to the perspective provided by the video in the standard interface where the user sees the video from the perspective of the robot's camera. The *close* perspective is zoomed out slightly and uses a virtual camera position somewhat above and behind the robot such that the front

Figure 6. Perspectives of the virtual 3D environment used in Experiment 4; clockwise from top left: 1st person, close, far, and elevated

half of the robot is also visible at the bottom of the screen. The *elevated* perspective zooms the map display out and places the camera behind and above the robot such that more of the map is visible in the interface. The *far* perspective zooms out further by placing the virtual camera position directly above the robot. It is far enough above the robot to allow the entire map to be visible on the screen. This is often referred to as a "bird's eye view." Figure 6 illustrates the different perspectives used for this experiment.

A maze environment was constructed on the first floor of the Museum of Idaho using cubicle wall dividers. On the second floor of the museum, a control station was set up that consisted of a laptop and monitor to display the interface and a joystick with which to control the robot. The participants could see the interface, but did not have the ability to see the actual robot or the maze itself as they drove the robot.

Prior to the experiment, each participant was instructed on the use of the joystick for controlling the robot. They were then requested to build a complete map of the maze as quickly as possible without running the robot into obstacles. Participants were also informed that the robot would prevent collisions, but that they should drive the robot in order to prevent such instances. Each participant used one of the four perspectives, which were assigned to volunteers in successive, cyclical order. Information, including the time required to complete the task, the initiative exercised by the robot, and the total joystick bandwidth used to guide the robot, was measured and recorded automatically and stored in a data file on the interface computer. Also, information on age, gender, and a self-assessment of video game skill (on a scale of 1 to 10) was recorded for each participant.

The results suggest that the 1^{st} person perspective was by far the most difficult to use, and the other three perspectives (close, elevated, and far) were similar to each other in their influence on the operator's ability to control the robot. In particular, participants using the 1^{st} person perspective took, on average, 133 seconds to discover the environment, while the close, elevated, and far perspectives had averages of 95, 96, and 97 seconds respectively. Additionally, participants using the 1^{st} person perspective had an average joystick bandwidth of 1,345, compared to 764, 724, and 693 for the close, elevated, and far perspectives respectively. There was not a significant difference in the number of times the robot took initiative to protect itself between any of the four different perspectives.

The results presented here suggest that the 1^{st} person perspective within the 3D display is inferior to the exocentric perspectives that show the robot and how it fits in relation to its environment. Although perspective is a critical factor in terms of time and joystick usage, it does not, at least for this study, seem to play a critical role in terms of operator navigational error (i.e., instances which necessitated robot initiative to protect itself). It is perhaps not surprising that perspective plays an important role; but what is surprising is that once the perspective moves from the 1st person to include the robot, there seems to be little difference between the various exocentric perspectives used. The close, elevated, and far perspectives all seemed to be very similar in terms of time, joystick usage, and robot initiative. This suggests that in comparison to the video module on the standard interface, the operator only needs a little more spatial information concerning obstacles near the robot in order to improve navigation significantly.

Additional studies will be necessary to further understand the benefits and limitations associated with different perspectives. Most likely, there will not be one optimal perspective. Rather, perspective should change based on the task element (e.g., navigation, search, patrol), the level of robot autonomy (e.g., direct human control, shared control, autonomous tasking), and the number of robots employed (Scholtz, 2002).

Conclusion

In this chapter we presented tools that improve a human-robot team performance in navigation and exploration tasks. The tools include behavior-based intelligence on the robot and a virtual 3D interface through which the operator views the information from the robot. The role of intelligence on the robot is to reduce the operator's need to understand the spatial environment immediately near the robot by empowering the robot to move and avoid obstacles without any operator control. The role of the virtual 3D interface is to improve the operator's ability to perceive and comprehend the spatial information around the robot, which enables the operator to issue more informed commands to the robot.

The reason the virtual 3D interface helps so much in the navigation and exploration experiments presented is because information is presented as it spatially relates to the robot. In contrast, the standard interface displays information in a manner that requires the operator to cognitively interpret the information into a holistic understanding of the robot's environment. This extra cognitive effort may impair the human's ability to anticipate or predict how the robot will respond to instructions. With the 3D interface, since information from the robot is automatically integrated by the manner of the presentation, the operator has more cognitive resources to anticipate how the robot will respond to instructions.

In the experiments where the human-robot interaction used the Shared control mode as opposed to the Safe control mode, performance also improved because the operator was not concerned with the low-level navigational control of the robot. Since the operator plays more of a supervisor role when the robot has navigational intelligence, the operator has more cognitive resources to allocate towards anticipating and predicting how the robot will respond to the environment.

By allowing the operator to visualize the robot's environment more clearly and providing the robot with intelligence to handle elementary aspects of navigation tasks, we bring the robot and the operator into a more unified frame of reference. With a unified reference frame the human and the robot move towards a true teaming paradigm where responsibilities and roles can shift dynamically depending on the needs of the human, the robot, or the task at hand. Improving the ability of robots and humans to work together has the potential to increase the applications and situations where robots can be effectively utilized as a valuable team member.

Acknowledgments

The authors wish to thank the following individuals for their contributions to this work: From the INL: Miles C. Walton, Julie Marble, PhD, Ron Boring, PhD, and Dan Henry; from Stanford Research Institute International: Kurt Konolige, PhD; from Brigham Young University: Mike Goodrich, PhD; and from Washington University in St. Louis: William D. Smart, PhD, and James Garner.

References

Baker, M., Casey, R., Keyes, B., & Yanco, H. A. (2004). Improved interfaces for human-robot interaction in urban search and rescue. In *Proceedings of the IEEE Conference on Systems, Man, and Cybernetics* (pp. 2960-2965).

Bruemmer, D. J., Few, D. A., Boring, R., Walton, M., Marble, J. L., Nielsen, C., & Garner, J. (2005). Shared understanding for collaborative control. *IEEE Transactions on Systems, Man, and Cybernetics Part-A, 35*(4), 494-504.

Burke, J. L., Murphy, R R., Coovert, M. D., & Riddle, D. L. (2004). Moonlight in Miami: A field study of human-robot interaction in the context of an urban search and rescue disaster response training exercise. *Human-Computer Interaction*, *19*, 85-116.

Burke, J. L., Murphy, R. R., Rogers E., Lumelsky, V. J., & Scholtz, J. (2004a). Final report for the DARPA/NSF interdisciplinary study on human-robot interaction. *IEEE Transactions on Systems, Man, and Cybernetics Part-C*, *34*(2),103-112.

Casper, J., & Murphy, R. R. (2003). Human-robot interactions during the robot-assisted urban search and rescue response at the World Trade Center. *IEEE Transactions on Systems, Man, and Cybernetics-Part B*, *33*(3), 367-385.

Clarke, D., Yen, S., Kondraske, G. V., Khoury, G. J., & Maxwell, K. J. (1991). *Telerobotic network workstation for system performance and operator workload monitoring* (Tech. Rep. No. 91-013R). National Aeronautical and Space Administration, Johnson Space Center.

Cooke, N. J., Salas, E., Cannon-Bowers, J. A., & Stout, R. (2000). Measuring team knowledge. *Human Factors, 42*, 151-173.

Dennett, D. (1981). True believers: The intentional strategy and why it works. In A. F. Heath (Ed.), *Scientific explanations: Papers based on Herbert Spencer lectures Given in the University of Oxford*, Reprinted in David Rosenthal (Ed.), *The Nature of Consciousness*, 1991.

Elfes, A. (1987). Sonar-based real-world mapping and navigation. *IEEE Journal of Robotics and Automation*, *3*(3), 249-265.

Endsley, M. R. (1988). Design and evaluation for situation awareness enhancement. In *Proceedings of the Human Factors Society 32nd Annual Meeting* (pp. 97-101).

Goodrich, M. A., Olsen Jr, D. R., Crandall, J. W., & Plamer, T. J. (2001). Experiments in adjustable autonomy. In *Proceedings of the IJCAI Workshop on Autonomy, Delegation, and Control: Interaction with Autonomous Agents*.

Khoury, G. J., & Kondraske, G. V. (1991). *Measurement and continuous monitoring of human workload associated with manual control devices* (Tech. Rep. No. 91-011R). National Aeronautical and Space Administration, Johnson Space Center.

Konolige, K. (2004). Large-scale map-making. In *Proceedings of the American Association for Artificial Intelligence* (pp. 457-463).

Gutman, J. S., & Konolige, K. (1999). Incremental mapping of large cyclic environments. In *Proceedings of the IEEE International Symposium on Computational Intelligence in Robotics and Automation (CIRA).*

Marble, J. L., Bruemmer, D. J., & Few, D. A. (2003). Lessons learned from usability tests with a collaborative cognitive workspace for human-robot teams. In *Proceedings of the Institute of Electrical and Electronics Engineers (IEEE) Conference on Systems, Man, and Cybernetics*.

Marble, J. L., Bruemmer, D. J., Few, D. A., & Dudenhoeffer, D. D. (2004). Evaluation of supervisory vs. peer-peer interaction for human-robot teams. In *Proceedings of the 37th Annual Hawaii International Conference on Systems Sciences*.

Moravec, H. P. (1988). Sensor fusion in certainty grids for mobile robots. *AI Magazine, 9*(2), 61-74.

Murphy, R. R. (2004). Human-robot interaction in rescue robotics. *Institute of Electrical and Electronics Engineers (IEEE) Transactions on Systems, Man, and Cybernetics-Part C*, *34*(2), 138-153.

Nielsen, C., Goodrich, M., & Crandall, J. (2003). Experiments in human-robot teams. In *Proceedings of the 2003 International Workshop on Multi-Robot Systems* (pp. 241-252).

Pacis, E. B., Everett, H. R., Farrington, N., & Bruemmer, D. J. (2004). Enhancing functionality and autonomy in man-portable robots. In *Proceedings of the SPIE Defense and Security Symposium*.

Pashler, H. (1990). Coordinate frame for symmetry detection and object recognition. *Journal of Experimental Psychology: Human Perception and Performance, 16*, 150-163.

Ricks, B. W., Nielsen, C. W., & Goodrich, M. A. (2004). Ecological displays for robot interaction: A new perspective. In *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*.

Scholtz, J. (2002). Human-robot interactions: Creating synergistic cyber forces. In *Proceedings of the 2002 Workshop on Multi-Robot Systems* (pp. 177-184).

Trouvain, B., Wolf, H. L., & Schneider, F. E. (2003). Impact of autonomy in multi-robot systems on teleoperation performance. In *Proceedings of the 2003 Workshop on Multi-Robot Systems* (pp. 253-264).

Woods, D., Tittle, J., Feil, M., & Roesler, A. (2004). Envisioning human-robot coordination in future operations. *Institute of Electrical and Electronics Engineers (IEEE) Transactions on Systems, Man & Cybernetics: Part-C: Special Issue on Human-Robot Interaction, 34*, 210-218.

Yanco, H. A., & Drury, J. (2004). Where am I? Acquiring situation awareness using a remote robot platform. In *Proceedings of the Institute of Electrical and Electronics Engineers (IEEE) Conference on Systems, Man and Cybernetics* (pp. 2835-2840).

Yanco, H. A., Drury, J. L., & Scholtz, J. (2004a). Beyond usability evaluation: Analysis of human-robot interaction at a major robotics competition. *Journal of Human-Computer Interaction, 19*, 117-149.

Yen, J., Yin, J., Ioerger, T., Miller, M., Xu, D., & Volz, R. (2001). CAST: Collaborative agents for simulating teamwork. In *Proceedings of the 17th International Joint Conference on Artificial Intelligence* (pp. 135-142).